

## I. THE CLAIMED INVENTION

The claimed invention (e.g., as recited in claim 1) is directed to a method (e.g., a computer-implemented method) for identifying relationships between text documents and structured variables pertaining to the text documents. The method includes providing a dictionary of keywords in the text documents, forming categories of the text documents using the dictionary and an automated algorithm, counting occurrences of the structured variables, the categories, and combinations of the structured variables and the categories for the text documents, and calculating probabilities of occurrences of the combinations of structured variables and categories.

Importantly, the method includes identifying a relationship between a structured variable and text documents included in a category based on a probability of occurrence of a combination of the structured variable and the category (Application at page 11, line 10-page 12, line 7; page 16, line 11-page 17, line 11; Figures 11 and 14; page 18, line 10-page 20, line 14).

Conventional methods of analyzing text documents cannot efficiently (e.g., automatically) identify interesting relationships between text documents (e.g., unstructured free-form text documents) and structured variables. Instead, words and phrases which frequently occur in the documents are plotted on a graph and users are required to determine for themselves whether an interesting relationship exists, which is labor intensive and time consuming (Application at page 1, line 17-page 2, line 1).

The claimed invention, on the other hand, identifies a relationship between a structured variable and text documents included in a category based on a probability of occurrence of a combination of the structured variable and the category (Application at page 11, line 10-page 12, line 7; page 16, line 11-page 17, line 11; Figures 11 and 14; page 18, line 10-page 20, line 14). Thus, unlike conventional methods, the claimed invention can efficiently (e.g., automatically) identify interesting relationships between the structured variables and categories of text documents (Application at page 11, lines 10-11; page 23, lines 1-8).

## II. THE ALLEGED PRIOR ART REFERENCES

### A. Lewak and Goldman

The Examiner alleges Lewak would have been combined with Goldman to form the invention of claims 1-2, 17, 23, 25 and 27-28. Applicant submits, however, that these

references would not have been combined and even if combined, the combination would not teach or suggest each and every element of the claimed invention.

Lewak discloses a method for accessing computer files. In the Lewak method, the user defines hybrid folders by describing the file contents of the files that belong to particular hybrid folders (Lewak at Abstract).

Goldman discloses method of using keyword extraction to examine a text collection of earthquake data (Goldman at Abstract).

Applicant respectfully submits that these references would not have been combined as alleged by the Examiner. Indeed, Lewak is directed to a method for accessing computer files, whereas Goldman is directed to a method of examining a text collection of data. Thus, these references are unrelated, and no person of ordinary skill in the art would have considered combining these disparate references, absent impermissible hindsight.

Further, these references clearly do not teach or suggest their combination. Therefore, Applicant respectfully submits that one of ordinary skill in the art would not have been so motivated to combine the references as alleged by the Examiner. Therefore, the Examiner has failed to make a prima facie case of obviousness.

Moreover, neither Lewak, nor Goldman, nor any alleged combination teaches or suggests a method for identifying relationships between text documents and structured variables pertaining to the text documents, which includes "*identifying a relationship between a structured variable and text documents included in a category based on a probability of occurrence of a combination of said structured variable and said category*", as recited, for example, in claims 1, 14, 17 and 23. As noted above, unlike conventional methods, the claimed invention can efficiently (e.g., automatically) identify interesting relationships between the structured variables and categories of text documents (Application at page 11, lines 10-11; page 23, lines 1-8).

Clearly, these novel features are not taught or suggested by the cited references or their combination. Indeed, the Examiner again expressly concedes that Lewak does not teach or suggest this feature on page 5 of the Office Action.

However, the Examiner alleges that the feature is taught by Goldman. The Examiner is incorrect.

The Examiner states on Page 14 of the Office Action that "Goldman performs knowledge discovery on a text database of Earthquake data looking for correlations between

earthquakes ... and time of day (structured variable) of the Earthquakes occurrence using statistical measures to determine significance of the combination of the two."

A careful reading of Goldman shows that the Examiner is **not correct** in his assertion. First of all, **there is no "structured variable" present in the text database of Earthquake data.** Instead, the authors simply "infer" time of day from mentions of keywords such as "morning", "evening", "afternoon", and "night" (e.g., see Goldman at Table 4, page 15). The weakness of this approach, as the authors themselves point out, is that these time indications are vague and not comparable, so morning might mean a wide range of times from midnight to Noon, while afternoon might encompass a much smaller time range. This makes statistics showing that morning to be the most frequent earthquake time very suspect.

The authors then go on to do a different analysis of actual structured data that is **not connected to the text.** That is, the authors take a new set of data that contains earthquakes and times and analyze it to show that time of day is not a factor. This analysis is done **separately from the text data analysis and does not directly use any of the categories or data entities used in the text analysis.** It is, therefore, not at all like the claimed invention which may use a data set (e.g., a single data set) consisting of both text and structured information.

Therefore, Goldman clearly does not teach or suggest identifying a relationship between a structured variable and text documents included in a category based on a probability of occurrence of **a combination of the structured variable and the category.**

Further, the "statistical measures" mentioned by the Examiner as being employed by Goldman are **entirely on the non-text data set.** Therefore, the "statistical measures" are not relevant to the statistical evaluation used in the claimed invention to determine if there is a significant text correlation between text and structured events (e.g., structured variables).

Therefore, Applicant submits that these references would not have been combined and even if combined, the combination would not teach or suggest each and every element of the claimed invention. Therefore, the Examiner is respectfully requested to withdraw this rejection.

#### **B. Goldszmidt**

The Examiner alleges Lewak would have been combined with Goldman and further combined with Goldszmidt to form the invention of claims 3-12, 14-16, 18-22 and 24. Applicant submits, however, that these references would not have been combined and even if

combined, the combination would not teach or suggest each and every element of the claimed invention.

Goldszmidt discloses a probabilistic approach to full-text document clustering which includes scoring document similarity based on probabilistic considerations. Similarity is scored according to the expectation of the same words appearing in two documents. The score enables the investigation of different smoothing methods for estimating the probability of a word appearing in a document, for purposes of clustering (Goldszmidt at Abstract).

Applicant respectfully submits that these references would not have been combined as alleged by the Examiner. Indeed, in contrast to Lewak and Goldman, Goldszmidt is directed to a method which estimates the probability of a word appearing in a document, for purposes of document clustering. Thus, these references are completely unrelated, and no person of ordinary skill in the art would have considered combining these disparate references, absent impermissible hindsight.

Further, these references clearly do not teach or suggest their combination. Therefore, Applicant respectfully submits that one of ordinary skill in the art would not have been so motivated to combine the references as alleged by the Examiner. Therefore, the Examiner has failed to make a prima facie case of obviousness.

Moreover, neither Lewak, nor Goldman, nor Goldszmidt, nor any alleged combination teaches or suggests a method for identifying relationships between text documents and structured variables pertaining to the text documents, which includes "*identifying a relationship between a structured variable and text documents included in a category based on a probability of occurrence of a combination of said structured variable and said category*", as recited, for example, in claims 1, 14, 17 and 23. As noted above, unlike conventional methods, the claimed invention can efficiently (e.g., automatically) identify interesting relationships between the structured variables and categories of text documents (Application at page 11, lines 10-11; page 23, lines 1-8).

Clearly, these novel features are not taught or suggested by the cited references or their combination. Indeed, the Examiner expressly concedes that Goldszmidt does not teach or suggest this feature on page 14 of the Office Action.

Therefore, Goldszmidt does not make up for the deficiencies of Lewak and Goldman.

Therefore, Applicant submits that these references would not have been combined and even if combined, the combination would not teach or suggest each and every element of the claimed invention. Therefore, the Examiner is respectfully requested to withdraw this

rejection.

### III. FORMAL MATTERS AND CONCLUSION


In view of the foregoing, Applicant submits that claims 1-12 and 14-28, all the claims presently pending in the application, are patentably distinct over the prior art of record and are in condition for allowance. The Examiner is respectfully requested to pass the above application to issue at the earliest possible time.

Should the Examiner find the application to be other than in condition for allowance, the Examiner is requested to contact the undersigned at the local telephone number listed below to discuss any other changes deemed necessary in a telephonic or personal interview.

The Commissioner is hereby authorized to charge any deficiency in fees or to credit any overpayment in fees to Assignee's Deposit Account No. 09-0441.

Respectfully Submitted,

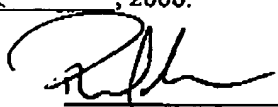
Date: 6/30/06

  
Phillip E. Miller, Esq.  
Registration No. 46,060

McGinn IP Law Group, PLLC  
8321 Old Courthouse Road, Suite 200  
Vienna, VA 22182-3817  
(703) 761-4100  
Customer No. 21254

### CERTIFICATE OF FACSIMILE TRANSMISSION

I hereby certify that the foregoing was filed by facsimile with the United States Patent and Trademark Office, Examiner James Blackwell, Group Art Unit #2176 at fax number 571-273-8300 this 30th day of June, 2006.

  
Phillip E. Miller  
Reg. No. 46,060